

Sequence match listing
(nucleic acid)

RESULT 4
ECNRFA
LOCUS ECNRFA 7320 bp DNA linear BCT 06-JUN-1995
DEFINITION E.coli nrfA gene.
ACCESSION X72298
VERSION X72298.1 GI:404302
KEYWORDS cytochrome c-552; nitrite reductase; nrfA gene.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 7320)
AUTHORS Darwin,A., Hussain,H., Griffiths,L., Grove,J., Sambongi,Y.,
Busby,S. and Cole,J.
TITLE Regulation and sequence of the structural gene for cytochrome c552
from Escherichia coli: not a hexahaem but a 50 kDa tetrahaem

JOURNAL	Medline	PubMed	Authors	Title	Journal	Remark	Reference	Comment	Source		
JOURNAL	MOI. Microbiol. 9 (6), 1255-1265 (1993)	95020657	2 (bases 1 to 7320)	Hussain, H.A.	Direct Submission	Submitted (28-MAY-1993)	H.A. Hussain, University of Birmingham, School of Biochemistry, Edgbaston, Birmingham, B15 2TT, UK	On Sep 30, 1993 this sequence version replaced g1:312132.	location/Qualifiers		
gene	1. 7320	/organism="Escherichia coli"	/db_xref="taxon:562"	/map="4367-4374 KOHARA MAP"	303.. 1739	/gene="nrfa"	303.. 1739	/gene="nrfa"	/transl_start=1		
gene	/protein_id="CAA51048.1"	/db_xref="GI:853826"	/db_xref="SWISS-PROT:P32050"	/translation="WRIKINARIFESLLLPFFFTSYHAQGTAPAKPVYAKNET PAPHPDQYLSWKATSEQSERVDALADPRLVIMAGTSPRYNKPGRHAPVTVR ETTLKAPKNAEDPILPMACWSCSPVAVAIQKDEGDYFHKRMARGEIVNLLGC ADCHNTASPEPAKPELTLSRPAARAMEAIQKPEKAGRFQOQSMVGCQCHVEYF DGKNCAYFPMWDGKMEKNEQYQYDKTAFSDWNTLSKTPMLKAQHPETWTAGTIG KNNTYKIDCHMPKYQNAEGKLYTDKRTGPNEDNACQACGCHKQDAALQVVAERKO SINDIKTIVEDQLVHAHEEAKALDAQATPAEKRPITDDIRHQAQWMDLAISGHYIM HAPEKDLMTGLTAMDKADARTKLARLARTGTHTHEIQIPDISTKERAAQATGLNNEQ IKAEPDQFIKTVIAPQWEQARKNGLLSQ"	1784.. 2350	/gene="nrfa"	1784.. 2350	/gene="nrfa"	/codon_start=1		
gene	/note="pentahaem C type cytochrome"	/codon_start=1	/transl_table=11	/protein_id="CAA51042.1"	/db_xref="GI:404303"	/translation="MSVLRSLTAGVILSGLLMSLNGITATPAQAQSDDRYEVTQNRN PDAALDCHKPDTEGMGKHSVIVPNPNKRLPVTCNGQSPQHRGSGVDVMEFNEDP MYKVEQNSVYCSCHLPOLQKAFPMWPHVHTKVAACASCHSLHPQDDMTQLSDKRI KTCVCHSDQRTNPNFNPASVPLKEQ"	2347.. 3018	/gene="nrfa"	2347.. 3018	/gene="nrfa"	/note="member of COOF family containing 4Fe /4S iron sulfur centre"
gene	/codon_start=1	/transl_table=11	/protein_id="CAA51043.1"	/db_xref="GI:404304"	/db_xref="SWISS-PROT:P32708"	/translation="MTWSRQFLITGVGLAANSGRAGRVAKTILNNGVRGYMDEES LCIGTACMDACREVNKPEGVSLTILRSEQGFPPVKTREFFKSCQCHDHAPCVIT VCTGASFRDAAAGIVDNPDLQVGCQYCIACPYRVRIRHPTVTADQCFQCRNTNIN QAKLACVACPTKALTEGNIIDDPNSEISQLRQKPTVRYKALGTAKPLRYVPEFKX GEVSQ"	3015.. 3971	/gene="nrfa"	3015.. 3971	/gene="nrfa"	/gene="nrfa"

[illegible]

Tue Nov 12 08:22:19 2002

us-09-902-

```

Db 950 TTACTTCGACGGCAAAAACAAAGCGGTTAAATTCCCGTGGGATGACGGCATGAAAGTCGA 1009
Qy 834 TCAGATCATGGCCTACTACGACGAGGACGGACACTCCGACTGGACGCACGAGCTCACGGG 893
Db 1010 AAATATGGAGCAGTATTACGACAAAATTGCCTTCTCTGACTGGACTAACTCCCTGTCGAA 1069
Qy 894 CGCCAAGGTGCTGAAGSCGCAGCACCCCGAGTTCGAGATGTACAACCAGGGCATCCACGC 953
Db 1070 AACGCCAATGCTGAAAGSCGCAGCACCCGGAATATGAAACCTGGACAGCGGGCATTACGG 1129
Qy 954 GAAGAGCGGCGTGGCCTGCGCGGACTGCCACATGCCGTTTCATGCGCGAGGGGGCGATGAA 1013
Db 1130 TAAAACACAGTGACCTGTATCGACTGCCATATGCCAAAAGTGCAGAACGCCGAAGGCAA 1189
Qy 1014 GGT---CAGCGACCACCAGGTGCGCAGCCCGCTGCTGAACATCAACCGCGCGTGCCAGAC 1070
Db 1190 ACTCTACACCGACCATAAAATTGGTAATCCGTTTGATAACTTCGCCCAGACTTGTGCCAA 1249
Qy 1071 GTGCCACAAGTGGAGCGAGGCGGAGCTGCTCCAGCGCGCGGAGACCATCCAGACGCGCAC 1130
Db 1250 CTGCCATACCCAGGACAAAGCTGCCTTGCAAAAAGTGGTCGCGGAACGTAAGCAGTCGAT 1309
Qy 1131 CTTCGAGACGCGGAACATCGCCATGGATGCGCTGGTGGACCTCATCCATGACATCGAGGC 1190
Db 1310 TAACGACCTGAAAATCAAGGTGAAAGATCAACTGGTTACGCTCACTTCGAAGCGAAAAGC 1369
Qy 1191 CGCGCAGAAGGCGGGGCGAGTCCGAGGAGGCGCTCGCCAAGGCGCGGACCTGCAGAAGCG 1250
Db 1370 AGCGCTGGATGCAGGCGCGACGGAAGCTGAAATGAAGCCAATTCAGGACGATATCCGTCA 1429
Qy 1251 CGCTCAATTCTACCTGGACTTCGTGGAGGCGGAGAACTCCATGGGCTTCCACGCGGACCA 1310
Db 1430 TGCCCAGTGGCGCTGGGATCTGGCGATCGCTTCCACGGCATTTCATATGCACGCACCGGA 1489
Qy 1311 GGAGGCGGTGCGCATCCTGAGCAACTCCAT 1340
Db 1490 AGAAGGTTTACGGATGCTCGGTACGCGAT 1519

```

Sequence match listing
(with amino acids)

RESULT 7
ECNRFA
LOCUS ECNRFA 7320 bp DNA linear BCT 06-JUN-1995
DEFINITION E.coli nrfA gene.
ACCESSION X72298
VERSION X72298.1 GI:404302
KEYWORDS cytochrome c-552; nitrite reductase; nrfA gene.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 7320)
AUTHORS Darwin,A., Hussain,H., Griffiths,L., Grove,J., Sambongi,Y.,
Busby,S. and Cole,J.
TITLE Regulation and sequence of the structural gene for cytochrome c552
from Escherichia coli: not a hexahaem but a 50 kDa tetrahaem
nitrite reductase
JOURNAL Mol. Microbiol. 9 (6), 1255-1265 (1993)
MEDLINE 95020657
PUBMED 7934939

REFERENCE 2 (bases 1 to 7320)
AUTHORS Hussain,H.A.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1993) H.A. Hussain, University of Birmingham,
School of Biochemistry, Edgbaston, Birmingham, B15 2TT, UK
REMARK revised by [3]

REFERENCE 3 (bases 1 to 7320)
AUTHORS Hussain,H.A.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-1993) H.A. Hussain, University of Birmingham,
School of Biochemistry, Edgbaston, Birmingham, B15 2TT, UK
COMMENT On Sep 30, 1993 this sequence version replaced gi:312132.

FEATURES
source Location/Qualifiers
1..7320
/organism="Escherichia coli"
/db_xref="taxon:562"
/map="4367-4374 KOHARA MAP"
gene 303..1739
/gene="nrfA"
CDS 303..1739
/gene="nrfA"
/codon_start=1
/transl_table=11
/protein_id="CAA51048.1"
/db_xref="GI:853826"

```

/db_xref="SWISS-PROT:P32050"
/translation="MTRIKINAKRIKIFSLIFEEFFTSVHAEDOTAPAKPVYVEAKNER
PAPHPPOIYSWKATISEQSERVADALAEIPRLIYIMAGYEPEDRNKPRGHAFATDNR
ETLRTGAPKNAEDGRLPMACWSCSPDVALRIGQDGEDGYEFGKMGARGPETVNNLC
ADCNNTASPEPAKPELTLRSRPAARAMEAIQGPKEAGFDOOSVCGCCHYEYF
DGKNKAVKPMDDGMKVENMEQYDYDKIAFSWNTSLSTKPLKAQHEYEYETWIGING
SINDKICIDCHMPKYONAEKGIYTHDKIGNPDPNAQTCANCHTODKALQVVARERO
HABEKGIMLTGTADKADARTKLARLATRGITHEIQIDIPDISKEKAQAQIGLMEQ
IKAEKODFIKTYIPQWEOARKNGLLSQ"
/gene="nrfb"
/length=1784
/genes="nrfb"
/note="penthaem C type cytochrome"
/codon_start=1
/transl_table=1
/protein_id="CA51042.1"
/db_xref="GI:404303"
/db_xref="SWISS-PROT:P32707"
/translation="MSVLRSLTAGVLAAGLSLNGITATPAASDSDREYTOORN
PAAACLDCHKPTEGHHGKHAIVINPNKLPYCTCHGQSPQHREGVKDWREYEP
MYKVEONSVCMSCHLPEOLOKAFMHQDVHVKVACASCHSLHPQODTMQTLSDKGR
KICVDCHSDORINPNPNASVPLKQOP"
/gene="nrfc"
/length=2347
/genes="nrfc"
/note="member of cooF family containing 4Fe /4S iron
sulfur centre"
/codon_start=1
/transl_table=1
/protein_id="CA51043.1"
/db_xref="GI:404304"
/db_xref="SWISS-PROT:P32708"
/translation="MTMSRROPFLTGCVLAAGVSTAGRVAKTLNINGVRYGMVDES
LCIGTACMDACREYKVPBGYSRLTILRSEPOGEPDVYKRFERKSCOHGHDACVD
VCPGASFRDASGLIVNPDLCVGCQYCIACPRVRFIHPVTAKDCPCRETNL
OAGKLPACVEACPTKALTFGNLDPNSEISOLRQKPYRYKRLALGTGPKLYRPFKY
GEVSQ"
/gene="nrfd"
/length=3015
/genes="nrfd"
/note="similar to NMDH ubiquinone oxidoreductase"
/codon_start=1
/transl_table=1
/protein_id="CA51044.1"
/db_xref="GI:404305"
/db_xref="SWISS-PROT:P32709"
/translation="MTQTSAFHFESESLWMDPIATYFLIGISAGLTVLAVLREYFQ
AGADSTLLRTTLTGAVGAILGILLVPHLTRPWTWKLMFHSFTSVSMGVMQLEQ
LYMVVLYMLAKIFEHDLALQORMLKLGIVOKVSLVTVPHHGETLMLVLAVLG
AYNGPLSALKSPFLNPLPLVPLEFSGISGAVALIAMIROSNPSTEFQFVH
RMEIPYWGTEIFLVAFFVGLAGDDKVALVAALGGGGMWFMVLGVAGGLIYPM
LKRPWNRSRSGITPAVLAAGSLVGLVLMRFLIYAGQLVA"
/gene="nrfE"
/length=4051
/genes="nrfE"
/note="homolog to ccl1 protein from Rhodobacter
capsulatus"
/codon_start=1
/transl_table=1
/protein_id="CA51045.1"
/db_xref="GI:581146"
/db_xref="SWISS-PROT:P32710"
/translation="MLPPLTAPAGVLRMPAMRLTICIGILAQFAVILLAPAVLYNCF
LISDFVLYVAOHSYSLSMELKLAVMGHEGSLIWLILSASALFAMRYROOTD
PLEPLPLVLSIMLAALLFLVYMSDPEVRIPPAIEGDLPMIOHPELIFHPCLY
LEGYGLVAVASALISLGEEDGACARSMMALPDSALAGIILGSMWVYCEGL
GLVVGSVENASLIPMLSATLLSHLSLTFRGHICHSLLAIVTMLSLGLTILY
RSGILVSHAFLADNVRVPLRSLFALLSLALYLGWRABDGBAVHPSGISRMLLI

```

```

LATLLFCVALLIVLVTLYPMIYGLMGRLSGAPYFNRAITLPEGLMLVVIYVAT
FVSGKRQOLDALVHAGVILFPAGVYVSVSRQESLINDPGQOVYTAGTFRECID
LOKANTSKATVIALEPDHQRIGELTPEPAFTEARQOMEPSIRMGJHDWYAVWG
EKGPRDPAVRLVYOSVVRWIMGGGLMIALGALLSGWRKKRDE"
/gene="nrfE"
/length=5702
/genes="nrfE"
/note="homolog to ccl2 protein from Rhodobacter
capsulatus"
/codon_start=1
/transl_table=1
/protein_id="CA51046.1"
/db_xref="GI:404307"
/db_xref="SWISS-PROT:P32711"
/translation="MNKGLTLLTLLFTCPAHQVVDYTMQFANPQOQOALNIAQLRC
POQONLLESNAPVAVSMRHOVYSVAGKKEVELIGMTERYGDGFVNPPLTGOT
LVLMALPVVLLLMALILMRVAKR"
/gene="nrfG"
/length=6082
/genes="nrfG"
/note="hydrophobic protein of unknown function"
/codon_start=1
/transl_table=1
/protein_id="CA51047.1"
/db_xref="GI:404308"
/db_xref="SWISS-PROT:P32712"
/translation="MKQPKIPVKMLTTLIMLVFCVGSYLLSPKQVAVRQORQ
PLHOFASOOTPEAOLOALODKIRAMPORSEBMALLEGYKONQYNSKLAVRQALQ
RGENAEYALATATVLYOASOHMTAOTAMIDKALALDSNETTALMLLASDAFQANY
AOATELMQKVDINSPRNRYQVIESIMANLQDRILD"
BASE COUNT 1621 a 1847 c 2071 g 1781 t
ORIGIN

Alignment Scores:
Pred. No.: 8.3e-31 Length: 7320
Score: 506.50 Matches: 135
Percent Similarity: 44.98% Conservative: 71
Best Local Similarity: 29.48% Mismatches: 152
Query Match: 19.44% Indels: 101
DB: 1 Gaps: 15

US-09-902-540-11926 (1-494) x EGNRPA (1-7320)
QY 43 GLVLAALysAsnProPheTYrArgValValGluLeuAspAspThrIleThrAspProGlu 62
Db 417 GAAGCGAAGAAAT-----GAA 431
QY 63 ValTrrpGlyArGAsnPhPheProLeuGlnITyRAspSerTYrLysArGthrValAspGln 82
Db 432 ACCTTTGGCCCGGCGACATCCCGATCAATATCTCTCGGAAGCCACCTGGAGACG--- 488
QY 83 ArgThrArGTyrGlyGlySerGluAlaValAlaArGthrProSerGlnAlaAspProArg 102
Db 489 -----TCAGACCGCGTT-----GAA 500
QY 103 ThrValValAlaGlnSerArGLeuGluAspProArgLeuValThrMetTrrpSerGly 122
Db 501 -----GAGCGCCCTGGCGGAGATCCACGCGTGTGATCCCTGTGGCGGG 545
QY 123 TyrAlaPheAlaThrAspPheArGLeuGluArgGlyHisAlaHisMetLeuAspAspGln 142
Db 546 TATCCCTTCTCGCGCATTAACAAGCGCGGACATGCTTTCTGCTGAGACGATG 605
QY 143 ValTrrpThrGluArgGlnHisValThrGlnGln-----ProGlyThr 156
Db 606 CGTGAACCCCTGGTACCGGTCGCGCGGAAGAACCTGGAAGATGGCTCGGTACCGATGCGA 665
QY 157 CysIleHisCysHisAlaSerValTyrValProTYrLysLeuGluYAspGlyAspLeu 176
Db 666 TCGTGAGTTGTTPAAAGCCG-----GATG 692

```

Tue Nov 12 08:22:12 2002

us-09-902-5

```
Qy 177 IleLysGlyPheGluLysMetAsnGlnMetProPheMetGluAlaArg----- 192
Db 693 GCGCGTCTGATCCAGAAAGACGGCGAAGATGGCTACTTCCACGGTAAATGGGCGCGCGC 752
Qy 193 -----LysLeuValGluHisProValSerCysIleAspCysHisAspProThrThr--- 209
Db 753 GGTCCGGAAATCGTCAACAAC---TTAGGTTGTGCCGATTGCCATAACACCGCCTCTCCA 809
Qy 210 -----MetGlnLeuArgValThrArgProGlyPheIleGluGlyIle 223
Db 810 GAGTTCGCCAAAGGCAAACCGGAGTTAACCTTTCCCGTCCGTATGCGGCTCGCGCGATG 869
Qy 224 AlaAlaLeuLysAlaSerGlnGlyValProAsnPheArgValAsnGlnAspAlaThrArg 243
Db 870 GAAGCCATT-----GGTAAACCT-----TTTGAGAAAGCCGGACGT 905
Qy 244 GlnGluMetArgThrTyrValCysGlyGlnCysHisValGluTyrTyrPheLysGlyLys 263
Db 906 TTCGACCAGCAATCGATGTTTGGCGTCAGTGCCATGTGGAGTATTACTTCGACGGCAAA 965
Qy 264 GluLysArgLeuThrTyrProTrpAlaLysGlyIleAsnIleAspGlnIleMetAlaTyr 283
Db 966 AACAAAGCGGTTAAATTCCTGGGATGACGGCATGAAAGTCGAAAATATGGAGCAGTAT 1025
Qy 284 TyrAspGluAspGlyHisSerAspTrpThrHisGluLeuThrGlyAlaLysValLeuLys 303
Db 1026 TACGACAAAATTGCCTTCTCTGACTGGACTAACTCCCTGTCGAAAACGCCAATGCTGAAA 1085
Qy 304 AlaGlnHisProGluPheGluMetTyrAsnGlnGlyIleHisAlaLysSerGlyValAla 323
Db 1086 GCGCAGCACC CGGAATATGAAACCTGGACAGCGGCATTACGGTAAAAACAACGTGACC 1145
Qy 324 CysAlaAspCysHisMetProPheMetArgGluGlyAlaMetLysVal---SerAspHis 342
Db 1146 TGTATCGACTGCCATATGCCAAAAGTGCAGAACGCCGAAGGCAAACTCTACACCGACCAT 1205
Qy 343 GlnValArgSerProLeuLeuAsnIleAsnArgAlaCysGlnThrCysHisLysTrpSer 362
Db 1206 AAAATTGGTAATCCGTTTGATAACTTCGCCAGACTTGTGCGAACTGCCATACCCAGGAC 1265
Qy 363 GluAlaGluLeu-----LeuGlnArgAlaGluThrIleGlnThrArgThrPhe 378
Db 1266 AAAGCTGCCTTGCAAAAAGTGGTCGCGGAACGTAAGCAGTCGATTACGACCTGAAAATC 1325
Qy 379 GluThrArgAsnIleAlaMetAspAlaLeuValAspLeuIleHisAspIleGluAlaAla 398
Db 1326 AAGGTTGAA-----GATCAACTGGTTCACGCTCACTTCGAAGCGAAAGCAGCG 1373
Qy 399 GlnLysAlaGlyGlnSerGluGluAlaLeuAlaLysAlaArgAspLeuGlnLysArgAla 418
Db 1374 CTGGATGCAGGCGCGACGGAAGCTGAAATGAAGCCAATTCAGGACGATATCCGTCATGCC 1433
Qy 419 GlnPheTyrLeuAspPheValGluAlaGluAsnSerMetGlyPheHisAlaAspGlnGlu 438
Db 1434 CAGTGGCGCTGGGATCTGGCGATCGCTTCCCACGGCATTATATGCACGACCGGAAGAA 1493
Qy 439 AlaValArgIleLeuSerAsnSerIleAsnPheSerArgLeuGlyGlnAsnAlaLeuArg 458
Db 1494 GGTTTACGGATGCTCGGT-----ACGCGGATGGAT 1523
Qy 459 ProSerGlyGlyAlaSerThrSerProThrThrArgProGlnGlyAlaProAla 476
Db 1524 AA-AGCGGCGGATGC-----ACGCACCAAACTGGCGCGCCTGCT 1561
```